MPRA

Munich Personal RePEc Archive

Testing Spatial Causality in Cross-section Data

Marcos Herrera Gómez and Manuel Ruiz Marín and Jesús Mur Lacambra

2014

Testing Spatial Causality in Cross-section Data

Marcos Herrera (CONICET-IELDE, National University of Salta); mherreragomez@gmail.com[1]

Manuel Ruiz (University of Murcia); manuel.ruizmarin@um.es

Jesús Mur (University of Zaragoza); jmur@unizar.es

Abstract

The paper shows a new non-parametric test, based on symbolic entropy, which permits detect spatial causality in cross-section data. The test is robust to the functional form of the relation and has a good behaviour in samples of medium to large size. We illustrate the use of test with the case of relationship between migration and unemployment, using data on 3,108 U.S. counties for the period 2003-2008.

[1]Corresponding author: IELDE, National University of Salta. Av. Bolivia 5150 (A4408FVY). Salta (Argentina).

1 Introduction

The study of causality has had a renewed interest in the last decades, especially in Economics as shown by the bibliometric study of Hoover (2004, p.4): 70% of the articles in the *JSTOR* archives, published during the period 1930-2001, contains words in a "*causal family ("cause", "causes", "causal", "causally" or "causality")*". The percentage increases up to 80% if the search is restricted to the field of econometric papers.

However, the impact of this debate has been surprisingly small in the field of spatial econometrics. For example, the subject index of the most influential textbook in this discipline, Anselin (1988), comprises approximately 500 items, not one of them being related to causation. The same can be said with respect to other popular textbooks, including the most recent of LeSage and Pace (2009): 1,000 headwords, none of which is related to the Hoover's causal family.

There are several reasons to explain the "*causality*" exception. Typical data in spatial econometrics, cross-sections in spatial models without time dynamics, does not facilitates the analysis. Indeed, the lack of adequate data increases the complexity of the analysis. Another argument is that cross-sectional models reflect general equilibrium solutions that must be interpreted in the long-term (Isard, 1971), so, there is no need to openly consider the issue of causality. This position is shared by part of the time series literature (i.e., Brockwell and Davis, 2003).

Our view is a bit more demanding: causality must be in the forefront of econometric modeling, irrespective of the type of data (time, space, mixed) or the nature of the relations. In particular, if this topic cannot be further evolved in spatial econometrics, the usefulness of this discipline will be limited to pure description, unable to corroborate or refute theories, which is clearly unsatisfactory.

Microeconomic studies in education, labor or health deal with questions such as "if x changes (i.e., professional training), what do you expect to happen with y (i.e., personal wage)?". To answer this question, Angrist and Pischke (2009) adopt the idea of natural experiments (that is, a change in a relevant policy that can be identified by non-statistical information) treating to replicate the experimentalist paradigm. It is assumed that causal variables can be manipulated or that we can create a situation *as if they were manipulable variables.* The aim is to mimic, as far as possible, the conditions of an experiment where we have two groups (i.e., treatment and control) before and after the intervention. Gibbons and Overman (2012) argue that the experimentalist approach is a good alternative to achieve identification, putting causality at the center stage in the debate of spatial econometrics[2]. The difficulty of this approach in a non-experimental field as spatial economics is obvious.

The dominant approach in Economics is non-experimentalist, sustained in the statistical properties of the process under study. Granger (1969) developed a notion of causality based on the *uni-directionality of time.* Temporal precedence and information content of the series are two major principles of this approach, which was supported also by Wiener (1956): "*For two simultaneously measured signals, if we can predict the first signal better by using the past information from the*

[2]The idea is not new. Among the first contributions in this field, we should mentioned the work of Isserman and Merrifield (1982).

second one than by using the information without it, then we call the second signal causal to the first one".

Recently, some researchers give the impression that spatial causality means: "... *when outcomes in area A affect outcomes in area B*," (Partridge *et al.*, 2012, p. 168). For example, Gibbons and Overman (2012) offer a questionable definition of spatial causality, using a spatial autoregressive model (*SAR*, in what follows):

$$y_i \quad = \quad \rho \mathbf{w}_i^{\prime} y + \beta x_i + u_i, \tag{1}$$

where i indexes locations, y_i is the outcome, x_i an explanatory variable, u_i an error term, ρ and β are parameters. The term $\mathbf{w}_i$ is a vector that captures "nearby" locations. According to the authors, parameter β captures the *causal effect* of x on y and ρ represents the *causal effect* of neighboring dependent variable. However, it is difficult to think in causal terms in relation to neighboring effects (*i.e.*, if this idea were translated to a time series context, it would be meaningless).

Our interpretation of causality in spatial models follows its time series analog: a process x, with a given spatial structure, causes a process y (with its own spatial structure) if the process x provides useful information about y. In the *SAR* model of equation (1), the spatial lag of the endogenous variable, $\mathbf{w}_i^{\prime} y$, captures the dependencies of that variable over the space, but this is not causality. Only x can cause y if it adds valuable information.

The objective of the paper is to introduce a new test for causality in pure cross-sectional spatial series, inspired on the Granger-Wiener approach. Our proposal is fully nonparametric and it is obtained without the explicit consideration of a model.

The paper is structured as follows. Section 2 provides a review of some of the main principles of Granger causality. Section 3 introduces the machinery of our approach. Section 4 formalizes our causality test for spatial cross-sectional data. Section 5 contains the results of the Monte Carlo experiment, while Section 6 presents a case study to the relation between unemployment rate and net migration using U.S. county data. Main conclusions appear in Section 7.

2 From Granger Causality to Spatial Causality

Granger test for causality, in combination with the test of Sims (1972), is by far the preferred option in applied econometrics. This literature has produced a simple, operational and testable definition to causality, based on three principles:

- Temporal precedence, in the sense that the effect should not precede the cause.
- Information completeness, in the sense that all information needed for the variables involved in the analysis is available.
- Temporal invariance that assures that causal mechanism remain constant throughout time.

The first introduces asymmetry so that, in binary relations (i.e., x causes y), is possible to discriminate between statistical association and causation. The second allows us to maintain uncertainty under control, and the third is vital for carrying out identification and statistical inference.

Under these premises, and assuming that the relevant series are weakly stationary (see, i.e., Hendry, 2004, for an extension to the non-stationary case), we say according to Granger (1969) that x is a causal variable for y if and only if $V\left(y_{t+1}|\Lambda_t\right) < V\left(y_{t+1}|\Lambda_t^{'}\right)$. Λ_t refers to the available information set at time t, which includes historical information up to period t of both series, x_t and y_t, together with z_t, the set of contextual variables, which create the conditions to study causality; $\Lambda_t^{'} \equiv \{y_{t-j}, z_{t-j}\}$, $\forall\, j \geq 0$, excludes past and present values of x_t. Finally $V(\cdot)$ is the forecasting error variance. In brief, x causes y if future values of y can be predicted more precisely if past values of x, in relation to y_{t+1}, are included in the data set.

Let us emphasize several important aspects related to this definition:

1. Granger causality does not require the specification of a direct causal mechanism (that is, we do not need a model). All that matters is observable predictive capacity (Davidson, 2000).

2. Investigators can never be sure whether an information set is complete. Statistical tests built on this way are condemned to be insufficient (Pearl, 2009).

3. Causality is a relation between variables that does not depend on their support, being time, space or whatever else (Davidson, 2000).

4. For a causal relation to be unambiguous, instantaneous and simultaneous causation should be excluded (Charemza and Deadman, 1997).

5. Finally, the existence of an attainable equilibrium for a pair of series requires (Granger) causality between them to provide the necessary dynamics (Maddala and Kim, 1999).

Granger causality is concerned with short-run forecastability so formal development of the causal relation between variables x and y is not needed. Theory is important, obviously, but this is not a theory-driven approach as, for example, Zellner (1988).

Contextual variables play a fundamental role: x causes y if it improves the forecasts of y after considering the general scenario where the relation is solved. We cannot be never sure that the set of contextual variables is, in fact, complete. This is the problem of *confounders,* or common causes, which lead to Pearl (2009, p. 195) to distinguish between necessary and sufficient conditions of causality: Granger causality is a necessary clause, not sufficient.

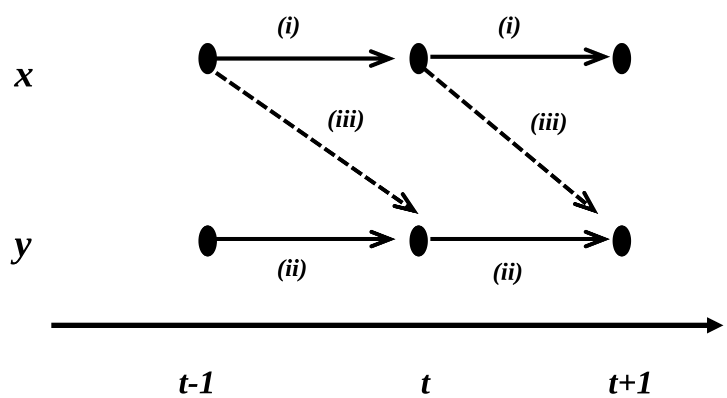
There must be a lag between the effect and the cause in order to undoubtedly identify the direction of causation. Instantaneous or simultaneous causality are hardly acceptable in the sense that they do not offer a clear and unambiguous description. The nature of the data is a different aspect. A cross-section corresponds to a single point in time where all the previous dynamics it is reflected. Obviously, if the time dimension disappears, a causality analysis based on temporal precedence will not be possible. The question is to have the capacity of separating causes from effects, even in these data sets.

Granger (1988) showed that if variables are cointegrated, their short-run dynamics can be expressed in the form of an Error Correction Mechanism and, thus, some variable must be caused by the others (or the whole set of variables may be caused by an external common factor). Cointegration is a relevant feature between variables and causality a necessary condition for cointegration.

Finally, it is important to remind that causality is a *relation between variables*, independently of their support. The spatial lag of a variable cannot cause the variable itself (the same as the time lag of y does not cause y); this is, as acknowledged by Partridge *et al.* (2012, p. 168), only spatial (auto)correlation. Space represents the domain from which data proceed, but is not a variable in a statistical sense.

The functioning of the Granger test for a bivariate system in a time domain can be represented graphically as in Figure 1.

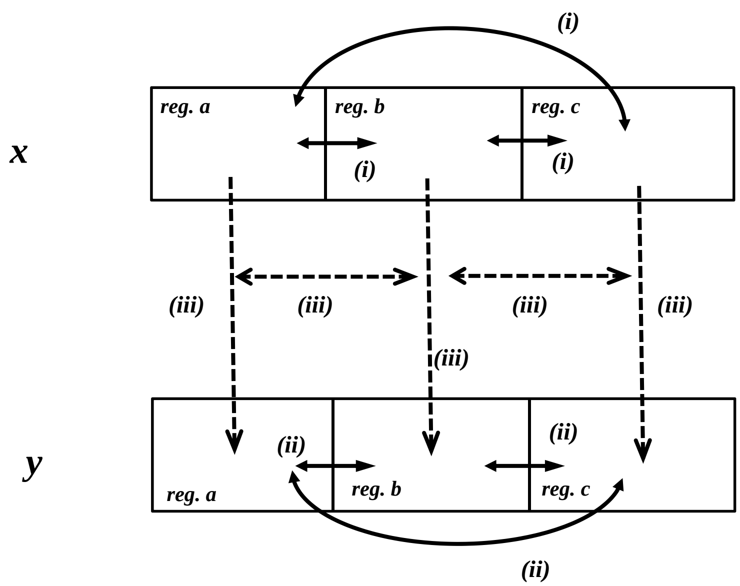
Figure 1: Representation of Granger Causality in a Time Domain.



Relations (i) and (ii) reflect serial dependence, relation (iii) adds causality from x to y. Time domain is naturally oriented which facilitates these flow-schemes.

The spatial domain, on the contrary, lacks of such natural orientation and the relations are, potentially, more complicated as in Figure 2. As before, solid lines (relations i and ii) indicate spatial dependence whereas causality is reflected by the broken lines. The differences between both figures are evident, especially because the lack of ordering in space. The concept of isotropy (Ripley, 1988) may introduce some order here although (i) usually it is related to a fundamental underlying factor that drives relations in space and (ii) isotropic processes are the exception rather than the norm in spatial economics. Our impression is that anisotropic relations are more interesting.

Figure 2: Representation of Granger Causality in a Spatial Domain.



3 Symbolic Dynamics and Entropy

Symbolic dynamics is based on the transformation of a series into a sequence of symbols, which captures relevant information of the series for statistical inference. In this sense, symbolic dynamics is a simplified description of a system dynamics (Hao and Zheng, 1998) for which we need a symbolization map. Then, once the series have been symbolized, we can evaluate the situation using different measures of entropy. The concept of informative content is clearly related to that of uncertainty (Bennett, 1998). The first section focuses on the symbolization procedure and the second on measures of entropy.

3.1 Symbolization Process

Let $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$ be two spatial processes of real data, where S is a set of locations. In order to symbolize the series, we have to define a non-empty finite set of symbols able to capture all the relevant information about the spatial process. This set is denoted by $\Gamma_n = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ whereas σ_i is the i-th symbol, for $i = 1, 2, \ldots, n$.

Symbolizing a process is therefore defining a map

$$f : \{x_s\}_{s\in S} \to \Gamma_n, \tag{2}$$

such that each element x_s is associated to a single symbol $f(x_s) = \sigma_i$. We say that location $s \in S$ is $\sigma_i - type$, relative to the series $\{x_s\}_{s\in S}$, if and only if $f(x_s) = \sigma_i$. We call f the *symbolization map*. The same process can be applied for the series y_s.

Let us introduce the bivariate process $\{Z_s\}_{s\in S}$ as:

$$Z_s = \{x_s, y_s\}. \tag{3}$$

For the bivariate process, we define the set of symbols Ω_n as the direct product of the two sets Γ_n; that is, $\Omega_n = \Gamma_n \times \Gamma_n$ whose elements are $\eta_{ij} = \left(\sigma_i^x, \sigma_j^y\right)$. The symbolization map for this process is:

$$g : \{Z_s\}_{s\in S} \to \Omega_n = \Gamma_n \times \Gamma_n, \tag{4}$$

where

$$g(Z_s) = (f(x_s), f(y_s)) = \eta_{ij} = \left(\sigma_i^x, \sigma_j^y\right). \tag{5}$$

We say that s is $\eta_{ij} - type$ for $Z = (x, y)$ or simply that s is $\eta_{ij} - type$, if and only if s is $\sigma_i^x - type$ for x and $\sigma_j^y - type$ for y.

The selection of the symbolization map is, essentially, a decision of the user. According to our experience the following procedure is simple and efficient for dealing with spatial data (see Matilla and Ruiz, 2008, 2009, López *et al.,* 2010, and Ruiz *et al.,* 2009 for other possibilities). For the series, i.e., $\{x_s\}_{s\in S}$ define the indicator function:

$$\tau_s = \begin{cases} 1 & \text{if} \quad x_s \geq M_e^x \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where M_e^x is the median of $\{x_s\}_{s\in S}$. Let $m \geq 2$ be the embedding dimension. For each $s \in S$, let N_s be the set formed by the $(m-1)$ neighbours of s. We use the term $m - surrounding$ to denote the set formed by each s and N_s, such that $m - surrounding$ $x_m(s) = \left(x_s, x_{s_1}, \ldots, x_{s_{m-1}}\right)$. We need to introduce a second indicator for each s_i, $i = 1, 2, \ldots, m-1$:

$$\iota_{ss_i} = \begin{cases} 0 & \text{if} \quad \tau_s \neq \tau_{s_i} \\ 1 & \text{otherwise} \end{cases}. \tag{7}$$

The symbolization map for $\{x_s\}_{s\in S}$ is simply:

$$f(x_s) = \sum_{i=1}^{m-1} \iota_{ss_i}. \tag{8}$$

Note that the cardinality of the set of symbols is m (that is, $m = n$ and $\Gamma_m = \{0, 1, \ldots, m-1\}$).

In short, the symbolization process consists of comparing, for each location s, the value τ_s with τ_{s_i} for each s_i in the set of $m-1$ nearest neighbours to s.

If this symbolization map is applied to an *i.i.d.* process, the probability of occurrence of each symbol is given by $p(\sigma) = C_\sigma^{m-1}/2^{(m-1)}$, where $C_\sigma^{m-1} = (m-1)!/[(m-1-\sigma)!\sigma!]$ denotes the combinations of $m-1$ elements taken $\sigma - in - \sigma$ for $\sigma \in \{0, \ldots, m-1\}$. For example, for $m = 4$, the expected relative frequencies for each symbol are: $p(\sigma = 0) = 1/8$, $p(\sigma = 1) = 3/8$, $p(\sigma = 2) = 3/8$, $p(\sigma = 3) = 1/8$.

3.2 Entropy: Definitions and Concepts

The entropy concept is at the very center of Information Theory, providing a measure of the uncertainty of a stochastic process. Let x be a discrete random variable that takes on values $\{x_1, x_2, \ldots, x_n\}$ with probabilities $p(x_i)$ for each $i = 1, 2, \ldots, n$, respectively.

Definition 1: The Shannon entropy, $h(x)$, of a discrete random variable x is defined as:

$$h(x) = -\sum_{i=1}^{n} p(x_i) \ln(p(x_i)).$$

Usually, when the base of the logarithm is equal to 2, the units are expressed in *bits*. We use the Neperian base, so the units are expressed in *nats*. Also, it is conventionally assumed that $0 \ln 0 = 0$; that is, adding terms equal to zero does not alter the entropy.

Definition 2: The entropy $h(x, y)$ of a pair of discrete random variables (x, y) with joint distribution $p(x, y)$, is:

$$h(x, y) = -\sum_{x}\sum_{y} p(x, y) \ln(p(x, y)).$$

Definition 3: Conditional entropy of variable x with respect to y is obtained as:

$$h(x|y) = -\sum_{x}\sum_{y} p(x, y) \ln(p(x|y)).$$

$h(x|y)$ measures the entropy of x that remains when y has been observed, assuming a joint distribution $p(x, y)$

Having symbolized the series, for a *embedding dimension* $m \geq 2$, it is easy to calculate the absolute and relative frequency for each symbol $\sigma_{i_s}^x \in \Gamma_n$ and $\sigma_{j_s}^y \in \Gamma_n$.

The absolute frequency of symbol σ_i^x is:

$$n_{\sigma_i^x} = \#\{s \in S | s \quad \text{is} \quad \sigma_i^x - type \quad \text{for} \quad x\}. \tag{9}$$

Similarly, for series $\{y_s\}_{s \in S}$, the absolute frequency of symbol σ_j^y is defined as

$$n_{\sigma_j^y} = \#\left\{s \in S | s \quad \text{is} \quad \sigma_j^y - type \quad for \quad y\right\}. \tag{10}$$

The relative frequencies can be estimated as:

$$p(\sigma_i^x) \equiv p_{\sigma_i^x} = \frac{\#\{s \in S | s \quad \text{is} \quad \sigma_i^x - type \quad for \quad x\}}{|S|} = \frac{n_{\sigma_i^x}}{|S|}, \tag{11}$$

$$p\left(\sigma_j^y\right) \equiv p_{\sigma_j^y} = \frac{\#\left\{s \in S | s \quad \text{is} \quad \sigma_j^y - type \quad for \quad y\right\}}{|S|} = \frac{n_{\sigma_j^y}}{|S|}, \tag{12}$$

where $|S|$ denotes the cardinality of set S; in general $|S| = N$. Similarly, we calculate the relative frequency for $\eta_{ij} \in \Omega_n$:

$$p(\eta_{ij}) \equiv p_{\eta_{ij}} = \frac{\#\{s \in S | s \quad \text{is} \quad \eta_{ij} - type\}}{|S|} = \frac{n_{\eta_{ij}}}{|S|}. \tag{13}$$

The *symbolic entropy* for a *two − dimensional* spatial series $\{Z_s\}_{s \in S}$ is the Shannon entropy

$$h_Z(m) = -\sum_{\eta \in \Omega_m^2} p(\eta) \ln(p(\eta)). \tag{14}$$

Similarly, we obtain the marginal symbolic entropies as

$$h_x(m) = -\sum_{\sigma^x \in \Gamma_m} p(\sigma^x) \ln(p(\sigma^x)), \tag{15}$$

$$h_y(m) = -\sum_{\sigma^y \in \Gamma_m} p(\sigma^y) \ln(p(\sigma^y)). \tag{16}$$

The symbolic entropy of y, conditioned by the occurrence of symbol σ^x in x is:

$$h_{y|\sigma^x}(m) = -\sum_{\sigma^y \in \Gamma_m} p(\sigma^y | \sigma^x) \ln(p(\sigma^y | \sigma^x)). \tag{17}$$

We also need the conditional symbolic entropy of y given x:

$$h_{y|x}(m) = -\sum_{\sigma^x \in \Gamma_m} \sum_{\sigma^y \in \Gamma_m} p(\sigma^x, \sigma^y) \ln(p(\sigma^y | \sigma^x)). \tag{18}$$

After a few manipulations, and using $p(\sigma^x, \sigma^y) = p(\sigma^x) p(\sigma^y | \sigma^x)$, we obtain:

$$h_{y|x}(m) = \sum_{\sigma^x \in \Gamma_m} p(\sigma^x) h_{y|\sigma^x}(m), \tag{19}$$

which means that the conditional symbolic entropy of y given x can be calculated as the average symbolic entropy of y conditioned by the symbolic realization of x.

4 Spatial Causality in Information

Before analyzing causality, a number of issues have to be considered to make sure that the analysis is consistent:

(1) The role of the space: if the variables were spatially independent, a traditional approach to causation it is preferable (e.g. Heckman, 1999, or Pearl, 2009).

(2) The relation between the variables: if the two variables are spatially independent, it does not make sense to talk about spatial causality.

(3) Assuming that (1) and (2) are satisfied, causality in information implies that there is a *one-way information flow* between the two variables.

This is the point that we address in continuation. We avoid (1), which amounts to the usual spatial dependence analysis (Lesage and Pace, 2009), and (2) which implies the use of some test of independence between spatial variables (Herrera *et al.*, 2013).

Let $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$ be two spatial processes and let $\mathcal{W}(x,y)$ be the set of spatial-dependence structures (that is, the set of spatial weighting matrices relevant for each variable) between x and y.

We use

$$\begin{aligned} \mathcal{X}_{\mathcal{W}} &= \{W_i x | W_i \in \mathcal{W}(x,y)\}, && (20) \\ \mathcal{Y}_{\mathcal{W}} &= \{W_i y | W_i \in \mathcal{W}(x,y)\}, && (21) \end{aligned}$$

to denote the sets of spatial lags of x and y given by all the weighting matrices in $\mathcal{W}(x,y)$.

Definition: *We say that $\{x_s\}_{s\in S}$ does not cause $\{y_s\}_{s\in S}$ under the spatial structures $\mathcal{X}_{\mathcal{W}}$ and $\mathcal{Y}_{\mathcal{W}}$ if*

$$h_{y|\mathcal{Y}_{\mathcal{W}}}(m) = h_{y|\mathcal{Y}_{\mathcal{W}},\mathcal{X}_{\mathcal{W}}}(m). \tag{22}$$

Then, we propose an unilateral non-parametric test for the following null hypothesis

$$H_0 : \{x_s\}_{s\in S} \textit{ does not cause } \{y_s\}_{s\in S} \textit{ under the spatial structure } \mathcal{X}_{\mathcal{W}} \textit{ and } \mathcal{Y}_{\mathcal{W}}, \tag{23}$$

with the following statistic:

$$\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) = \hat{h}_{y|\mathcal{Y}_{\mathcal{W}}}(m) - \hat{h}_{y|\mathcal{Y}_{\mathcal{W}},\mathcal{X}_{\mathcal{W}}}(m). \tag{24}$$

That is, if $\mathcal{X}_{\mathcal{W}}$ does not contain extra information about y then $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$=0, otherwise, $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) > 0$. The alternative is that the null hypothesis of (23) is not true.

In order to remain in a model-free framework, it is preferable to determine the significance of the test using bootstrap methods. Our approach follows the guidelines of *non-overlapping time block bootstrap* of Carlstein (1986). We design a *spatial block bootstrap* (SBB; Appendix A) to break down the dependence between the series but preserving most of the spatial structure in each series. The SBB procedure, with a number B of bootstraps, consists of the following steps:

1. Compute the value of the statistic $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ using the original data, $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$.
2. Divide each spatial series into $b = N/l$ contiguous observational blocks of l units. Remember that N is the sample size. By contiguous observational blocks we mean that the observations of each block are contiguous according to the W matrix. The blocks cannot overlap and must cover the entire space.

3. Generate two new samples of length N by resampling, with replacement, the b blocks of x and y. Let us call $\{x_s(i)\}_{s\in S}$ and $\{y_s(i)\}_{s\in S}$ the two bootstrapped series, where i is the number of the bootstrap sample.

4. Estimate the bootstrapped realization of the statistic $\hat{\delta}^{(i)}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ using the bootstrapped series $\{x_s(i)\}_{s\in S}$ and $\{y_s(i)\}_{s\in S}$.

5. Repeat $B-1$ times steps 3 and 4 to obtain B bootstrapped realizations of the statistic $\left\{\hat{\delta}^{(i)}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right\}_{i=1}^{B}$.

6. Compute the estimated bootstrap $p-value$:

$$p_{boots} - value\left(\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right) = \frac{1}{B}\sum_{i=1}^{B}\tau\left(\hat{\delta}^{i}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) > \hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right) \tag{25}$$

   where $\tau(\cdot)$ is an indicator function that assigns 1 if inequality is true and 0 otherwise.

7. Reject the null hypothesis that $\{x_s\}_{s\in S}$ does not cause $\{y_s\}_{s\in S}$ under the spatial structure $\mathcal{W}(x,y)$ if

$$p_{boots} - value\left(\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right) < \alpha \tag{26}$$

   for a nominal size α.[3]

5 Monte Carlo Simulations

The objective of this section is to study the behavior of the test in (24) for finite samples. We examine the empirical size not only when x and y are *i.i.d.*, but also if they are spatially dependent without spatial causality. We also want to examine the power of the test in presence of linear and non-linear spatial causality.

5.1 Experimental Design

Each experiment starts by obtaining a random map in a bivariate system of coordinates. Then a normalized W matrix is built following the $m-1$ nearest neighbours criterion

The following global parameters are involved in the *D.G.P.*:

$$N \in \{320, 560, 800\}, \; m \in \{4, 5, 6\}, \tag{27}$$

where N is the sample size and m is the *embedding dimension*. We simulate linear and non-linear relations between the two variables x and y:

DGP1

$$y = (I - \rho W)^{-1}(\beta x + \theta W x + \varepsilon), \tag{28}$$

[3] A Matlab routine for this estimation procedure can be downloaded for free from the authors' Web site: https://sites.google.com/site/spatialcausality/codes

DGP2

$$y = \exp\left[(I-\rho W)^{-1}(\beta x + \theta W x + \varepsilon)\right], \tag{29}$$

where $x \sim \mathcal{N}(0,1)$, $\varepsilon \sim \mathcal{N}(0,1)$, $cov(x,\varepsilon) = 0$ and $\rho \in \{0.0; 0.3; 0.5; 0.7\}$.

We are very interested in controlling the intensity of the signal between the two variables using the expected $R^2_{y/x}$ coefficient of a linear equation like:

$$y = \beta x + \theta W x + \varepsilon. \tag{30}$$

The expected $R^2_{y/x}$ in (30), under the assumptions above, can be expressed as:

$$R^2_{y/x} = \frac{\beta^2 + \left(\theta^2/m-1\right)}{\beta^2 + \left(\theta^2/m-1\right) + 1}.$$

We have considered two values for this coefficient:

$$R^2_{y/x} \in \{0.6; 0.8\}. \tag{31}$$

For simplicity, in all cases we use $\beta = 0.5$. The spatial lag parameter of x is obtained simply as: $\theta = \sqrt{\frac{(1-m)(\beta^2(1-R^2)-R^2)}{1-R^2}}$.

The empirical size has been estimated using two independent processes such as:

$$y = \rho_y W y + \varepsilon_y, \tag{32}$$

$$x = \rho_x W x + \varepsilon_x, \tag{33}$$

where as before $\varepsilon_j \sim \mathcal{N}(0,1)$; $j = x, y$ and:

$$\rho_y \in \{0.0; 0.4; 0.8\},$$

$$\rho_x \in \{0.0; 0.3; 0.7\}.$$

As a general rule, following Rohatgi (1976), we consider that there should be, on average, *5* observations for each symbol whose frequency is estimated (this decision affects the sample size and the number of variable used in the experiment). Finally, in all cases, we consider that the $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ test contains the full information set: $\mathcal{Y}_{\mathcal{W}} = \{Wy\}$, $\mathcal{X}_{\mathcal{W}} = \{Wx\}$.

5.2 Performance for Finite Samples

The size is the percentage of false rejections of the null hypothesis of no causality from x to y. Under the best of circumstances, this empirical size should be closed to the nominal level. If the empirical size is smaller than the nominal level, the test is called conservative. If the empirical size is larger than the nominal level, the test is called oversized.

It is important to stress that all the results below have been obtained after a double testing sequence. That is, *Global empirical size* reflects the percentage of cases in which the hypothesis of (23) is incorrectly rejected and *simultaneous* its complement (y does not cause x) are correctly no-rejected. That is, we compute all false positive of unique direction from x to y. Similarly, *Global estimated power* is the percentage of rejections of the null of (23) with the *simultaneous* no-rejection of the complementary hypothesis (y does not cause x). Simultaneous rejection of both null hypothesis, as indicated before, does not allow identifying an unambiguous direction of causality. These cases are excluded when computing the estimated power.

Table 1 shows the empirical size of the statistic at a 5% nominal level. As it is evident from this table, the best situation for the $\hat{\delta}\left(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}\right)$ test corresponds to *i.i.d.* series (ρ_y=$\rho_x = 0$). The test becomes conservative for higher values in any of the two parameters of spatial dependence.

Table 1: Global empirical size of $\hat{\delta}\left(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}\right)$ Test at 5% level.

ρ_y	ρ_x	$N = 320$	$N = 560$		$N = 800$		
		$m = 4$	$m = 4$	$m = 5$	$m = 4$	$m = 5$	$m = 6$
0.0	0.0	5.3	4.7	4.3	4.9	5.4	4.0
	0.3	4.7	5.4	3.0	4.4	3.1	2.8
	0.7	1.5	2.7	2.0	3.6	2.1	0.5
0.4	0.0	4.1	3.8	2.8	5.4	3.8	3.1
	0.3	2.7	4.6	2.4	4.0	3.1	2.1
	0.7	0.5	2.6	0.6	2.8	1.3	0.1
0.8	0.0	1.5	2.5	0.7	3.0	1.2	0.5
	0.3	0.8	1.4	0.5	2.2	0.6	0.2
	0.7	0.2	0.7	0.1	1.1	0.1	0.1

Note: Boots: 399. Blocks: 8. Replications: 1000.

Table 2: Global estimated power of $\hat{\delta}\left(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}\right)$ Test at 5% level. DGP 1: Linear Case

N	ρ	$R^2_{y/x} = 0.6$			$R^2_{y/x} = 0.8$		
		$m = 4$	$m = 5$	$m = 6$	$m = 4$	$m = 5$	$m = 6$
320	0.3	94.4	–	–	91.6	–	–
	0.5	98.5	–	–	97.8	–	–
	0.7	92.0	–	–	95.8	–	–
560	0.3	75.2	98.2	–	69.1	97.1	–
	0.5	91.4	99.0	–	89.4	98.9	–
	0.7	97.9	99.9	–	97.1	99.7	–
800	0.3	62.5	93.0	100.0	53.4	89.7	99.9
	0.5	84.0	98.8	100.0	79.9	98.4	100.0
	0.7	96.4	99.9	100.0	95.9	99.6	100.0

Note: Boots: 399. Blocks: 8. Replications: 1000. Empty cells correspond to cases where number of obs. per symbol is below of 5.

Table 2 shows the results obtained for the linear case, $DGP1$. As expected, the estimated power increases with sample size and embbeding dimension. For high values of the coefficient of spatial autocorrelation of y, such as 0.7, and small sample sizes the power slightly decreases. This estimated power is *100%* in many cases.

For non-linear process, Table 3, the test shows a good performance in almost all simulated cases. The estimated power improves as the sample size increases, the same as the linear case.

Table 3: Global estimated power of $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ Test at 5% level. DGP 2: Non-Linear Case

N	ρ	$R^2_{y/x} = 0.6$			$R^2_{y/x} = 0.8$		
		$m = 4$	$m = 5$	$m = 6$	$m = 4$	$m = 5$	$m = 6$
320	0.3	93.0	–	–	99.5	–	–
	0.5	97.3	–	–	99.7	–	–
	0.7	91.6	–	–	80.2	–	–
560	0.3	79.8	95.1	–	71.7	92.0	–
	0.5	88.8	97.2	–	86.9	96.5	–
	0.7	96.5	99.5	–	94.2	99.4	–
800	0.3	69.3	86.6	98.4	67.9	79.8	96.4
	0.5	86.6	94.5	99.4	79.5	93.1	99.2
	0.7	93.3	98.3	99.9	91.1	98.0	99.7

Note: Boots: 399. Blocks: 8. Replications: 1000. Empty cells correspond to cases where number of obs. per symbol is below of 5.

Overall, these results are quite satisfactory in spite of the test being conservative.

6 Unemployment and Migration

There is a huge literature in labour economics regarding the relationship between migration and unemployment. At the risk of simplifying, we can say that the conclusion that immigration causes unemployment is part of the neoclassical paradigm (Borts and Stein, 1964). Assuming homogeneity in the labor force and perfect competition in the market of goods, workers move to prosperous regions attracted by higher salaries. The inflow increases labor supply in these regions (direct effect). In turn, migrants increase the consumption of local goods, leading companies to hire more workers (indirect effect). It is customary to assume that direct effects prevail over the indirect effects, raising unemployment in the regions of destination.

New Economic Geography (Krugman, 1991) assumes imperfect competition on the goods market and rigid labor markets but the conclusion is the same: migration causes unemployment. Migrant flows stimulate agglomeration economies in the regions of destination, increasing inequalities between the central and peripheral region, unemployment included.

Other strand of literature concludes that unemployment is the cause of migration. Pissarides and Wadsworth (1989) argued that people move from places where they are not fully employed to places offering greater possibilities of being employed. In fact as it is well documented, unemployment in the place of origin increases the willingness to migrate (Antolin and Bover, 1997).

Section 6.1 reviews some of the applied literature on the subject. The conclusion is that there is abundant evidence supporting the two hypotheses. Section 6.2 applies our methodology to the case of U.S. counties for 2003-2008.

6.1 Literature Review

The debate between unemployment and migration has a long tradition in the economic literature, where we can find many different studies. From our perspective, this studies can be divided into two groups. The first considers the simultaneity between the variables and aims to test causality, usually according to a Granger's approach. The second group assumes a certain previous position, avoiding the causality problem.

Table 4 presents a summary of recent causal studies.

Table 4: Causality Studies between Unemployment and Migration.A summary.

Authors	Country/Data	Main Variables	Methods	Main Conclusion
Pope and Withers (1993)	Australia/Annual 1861-1991	Unemp.; Immigr. and 5 other variables.	VAR model	Unemp. causes Immigr.
Marr and Siklos (1994)	Canada/Quaterly 1962:4-1985:4	Unemp.; Immigr. and 3 other variables.	Bivariate relationship with control variables	Unemp. causes Immigr. before 1978. Immigr. causes Unemp. after 1978.
Marr and Siklos (1995)	Canada/Annual 1926-1992	Unemp.; Immigr; Wage; GDP	Bivariate relationship; VAR model	Immigr. causes Unemp.
Tian and Shan (1999)	Australia/Quarterly 1983:3-1995:4	Unemp.; Net Immigr.; and 4 other variables.	VAR model	No evidence of causality
Konya (2000)	Australia/Quarterly 1981:2-1998:4	Unemp. and Net Migr.	VAR model	Net Migr. causes Unemp.
Feridun (2004)	Finland/Annual 1981-2001	Unemp., Foreign Immigr. and GDP	VAR model	Immig causes Unemp.
Gross (2004)	Canada/Annual 1980-1995	Unemp., Immigr, 2 other variables and 6 exoge. variables	VECM model	Immigr. causes Unemp. in the short term. No evidence in the long term
Feridun (2005)	Norway/Annual 1983-2003	Unemp., Foreign Immigr. and GDP	VECM model	No evidence of causality
Feridun (2007)	Sweden/Annual 1980-2004	Unemp., Foreign Immigr. and GDP	VECM model	Unemp. causes Immigr.
Islam (2007)	Canada/Quarterly 1961:1-2001:1	Unemp., Immigr. Wage and GDP	VECM model	Unemp. causes Immigr. in the short term. No evidence in the long term
Basile et al. (2012)	Italy/Annual 1995-2006	Unemp., Net Migr.	Spatial Dynamic Panel Data model	Net Migr. causes Unemp.

Note: VAR: Vector Autoregressive, VECM: Vectot Error Correction Model.

Table 5 completes the information with a collection of non-causal studies. It is clear that the evidence is also very dispersed in this case.

Table 5: Non-Causal Studies between Unemployment and Migration. A summary.

Authors	Country/Data	Main Variables	Methods	Main Conclusion
Da Vanzo (1978)	US/Annual 1971-1972	Migr. (change of residence), Unemp, and other control variables	OLS-Probit models	Unemp. affects the probability to Migr.
Pissarides and Wadsworth (1989)	Great Britain/Annual 1976/7 and 1983/4	Migr. (change of residence), Unemp,and other control variables	Logit regression model	Unemp. affects the probability to Migr.
Pissarides and McMaster (1990)	Great Britain/Annual 1961-1982	Regional Unemp., Net migr., wage, control variables	Pooled regression model	Unemp. affects the Net migr.
Blanchard and Katz (1992)	US/Annual 1972-1990	Unemp., Net Migr., Wages, and other control variables	Structural Vector Autoregressive model	Net Migr. has a crucial impact in the convergence of regional unemp.
Antolin and Bover (1997)	Spain/Annual 1987-1991	Differential regional unemp., Migr.(change of residence), and other control variables	Individual Pooled Regression, Logistic model	Unemp. has a strong impact on Migr. using individual data. No evidence using regional agreggates
Faini et al. (1997)	Italy/ Survey 1995	Internal migr. (willing to migrate), regional unemp., and other control variables	Multinomial logit model	Unemp. has a strong impact on Migration
Pischke and Velling (1997)	Germany/Annual 1985-1989	Unemp., Foreign Immigr., and other control variables	OLS-IV model	Foreign Inmigr. has a small effect on Unemployment.
Basile and Causi (2007)	Italy/Annual 1991-2000	Net migr., Regional Unemp.,and other control variables	Seemingly Unrelated Regression	Unemp. has a significant effect on Net Migr. after 1995

Note: OLS: Ordinary Least Squares, IV: Instrumental Variables.

Our conclusions after this revision are somewhat mixed: (i) There is no consensus about the relation between unemployment and migration; (ii) The dominant approach uses national, or regional, time series data; (iii) Granger notion is the preferred option in causality studies.

6.2 Net Migration and Regional Unemployment in US Counties

This section analyses the relation between unemployment and net migration in 3,108 U.S. counties for the period 2003-2008. The purpose is testing for causality between the two variables and, if so, detecting the direction of causation using the methodology introduced previously.

We use annual data. The unemployment rate is equal to the number of unemployed divided by the labour force[4]. Net migration is the difference between in-migration to an area and out-migration from the same area, on a July-to-July comparison, as a proportion of an area's population at the midpoint of the time period and expressed per 1,000 population[5].

Table 6 presents a summary for the 3,108 US counties during the six years. The unemployment rate fell for the whole period 2003-2007. The American rate (county level) was 5.97% in 2003, falling to 4.84% in 2007. Net migration grew in the first four periods, from an average of 0.85% in 2003 to 2.26% in 2006. Years 2007 and 2008 show a decrease in the aggregated value reaching a negative mean to -0.14%. The dispersion differs among variables and between periods. In the case of unemployment, the standard deviation fell from 2003 to 2006 and then increased over 2008 reaching a value of 2.07. Dispersion in the net migratory rate grew throughout the first four periods but decreased for the last two years.

Table 6: Unemployment rate and net migration by US county

	Unemployment Rate (%)		Net Migration (‰)	
Year	Average	Stand. Dev.	Average	Stand. Dev.
2003	5.97	1.93	0.85	13.83
2004	5.63	1.77	1.63	15.34
2005	5.38	1.78	1.28	15.39
2006	4.89	1.67	2.26	18.44
2007	4.84	1.69	1.15	14.77
2008	5.76	2.07	−0.14	14.42

Source: *US Census, Bureau of Labor Statistics.*

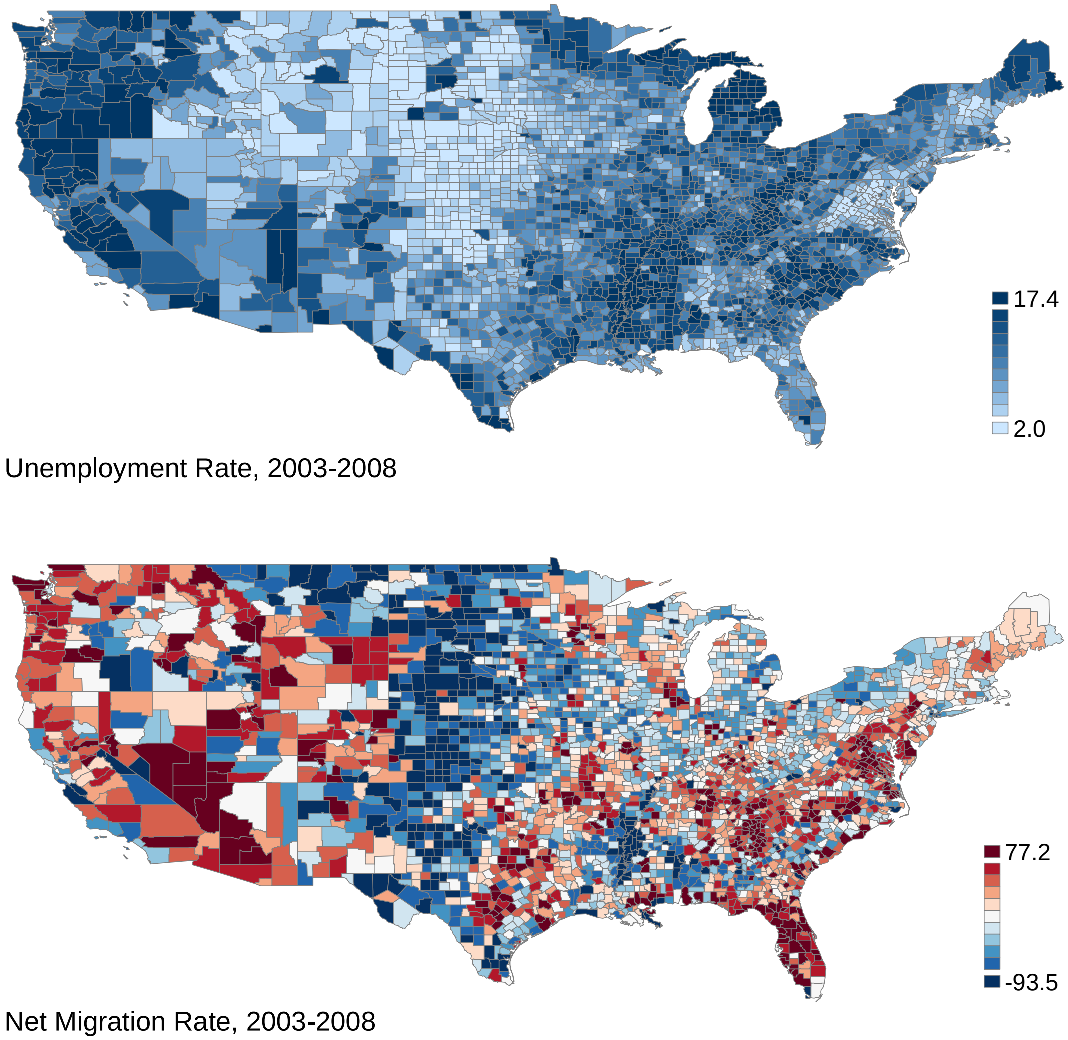
Net migration is measured from July to July but unemployment is measured on a calendar year. In order to reduce the disagreement, we use averaged data for 3-years periods: 2003-2005, 2004-2006, 2005-2007, 2006-2008, plus the average corresponding to the overall period 2003-2008.

In Figure 3, we present the spatial distribution of unemployment and net migration (averaging all years). There are clusters of high unemployment, in the first map, in West Coast counties and in some Northeastern regions such as Michigan, Wisconsin and Maine. Furthermore, we can detect clusters of low unemployment rate in North Central regions and in some states of East Coast (Virginia, Maryland, Columbia, Vermont and New Hampshire).

[4]Unemployment rate was estimate using labor force data by county, annual averages (Bureau of Labor Statistics, Local Area Unemployment Statistics program)

[5]Includes domestic and international migration (United States Census Bureau, Demographic Components of County Population Change).

Figure 3: Spatial Distribution of Unemployment and Net Migration by County. Average 2003-2008.



With regards to migratory flows, the Western and Eastern states are recipients whereas the central states are generators of migrants.

As said, there are 3,108 counties in the sample. Applying the rule of $m^3 \cdot 5 \cong N$, we choose an embedding dimension m equal to 9, which means that for each county we are using the 8 nearest neighbours to construct the symbol.

Before testing for causality, we tested spatial dependence using the Moran's I. The dependence between the two variables, taking into account the spatial structure, has been tested by means of the ψ_2 test whose null hypothesis is that the two variables are spatially independent (Herrera *et al.*, 2013). The contact matrix corresponds to the row-standardized version of the 8 nearest neighbours. Main results of these preliminary stages appear in Table 7.

Table 7: Spatial Dependence Tests

Year	*Unemp.* Moran's I	*Net Mig.* Moran's I	*Net Mig. − Unemp* $p-value\,(\psi_2)$
2003-2005	0.55***	0.40***	0.030**
2004-2006	0.56***	0.39***	0.047**
2005-2007	0.58***	0.38***	0.040**
2006-2008	0.61***	0.36***	0.027**
2003-2008	0.58***	0.42***	0.045**

Note: * : signif. 10%, ** : signif. 5%, *** : signif. 1%.
Permuts: 499, W : 8 nearest neighs.

Table 8 shows the results of spatial causality test introduced in Section 4. We detect spatial causality from net migration to unemployment. Indeed, there is a clear signal of causality in our county data set in the sense that the information flow is unidirectional from migration to unemployment.

Table 8: Results of Spatial Causality Test

H_0	$Unemp. \nRightarrow Migr.$	$Migr. \nRightarrow Unemp.$	*Conclusion*
Periods	$p-value$	$p-value$	
2003-2005	0.100	0.045	$Migr. \Rightarrow Unemp.$
2004-2006	0.882	0.048	$Migr. \Rightarrow Unemp.$
2005-2007	0.827	0.049	$Migr. \Rightarrow Unemp.$
2006-2008	0.812	0.047	$Migr. \Rightarrow Unemp.$
2003-2008	0.652	0.050	$Migr. \Rightarrow Unemp.$

Note: "$\nRightarrow$" means *does not cause* and "$\Rightarrow$" means *causes*. $m = 9$, Boots: 399, Blocks: 42 (74 obs. by block).

In sum, space is relevant to interpret the relation between net migration and unemployment by counties, the two variables are not spatially independent and there exists causality from the first variable to the second.

7 Summary

Pagan (1989, p. 89) admitted his disappointment with the notion of causality in the field of Econometrics, writing that *"there was a lot of high powered analysis of this topic, but I came away from a reading of it with the feeling that it was one of the most unfortunate turnings for econometricians in the last two decades, and it has probably generated more nonsense results than anything else during that time"*. We partially agree with him. Indeed we think that causality is one of the major concepts developed in modern economics and, because of its vital importance, we need of powerful techniques to properly deal with it.

Following the Granger-Wiener tradition, we identify causality with the principle of *incremental informative content*. Intuitively, our definition establishes that the cause variable should provide additional and unique information about the effect variable.

The test of spatial causality compares two measures of conditional entropy. The first measure uses all the information available in the space concerning the effect variable; the second adds the spatial information of the variable assumed to be the cause. The test is intuitive, easy to obtain and does not need of any hypothesis about functional form, distribution function, or other aspects of the specification. It is a fully non-parametric causality test.

Our proposal shows a good behavior in samples of medium to large sizes, as evidenced in the Monte Carlo. Moreover, it is robust to the functional form of the relation, linear or non-linear.

We have applied our methodology to the debate about unemployment versus migration. There is an abundant literature supporting the assumptions that unemployment causes migration and that migration causes unemployment. In our case, using data on U.S. counties for the 2003-2008 period, we have found clear evidence supporting the first hypothesis: migration causes unemployment.

References

[1] Angrist, J. and J. Pischke (2009): *Mostly harmless econometrics: An empiricists companion.* Princeton University Press.

[2] Anselin, L. (1988): *Spatial econometrics: Methods and models.* Dordrecht, Kluwer.

[3] Antolin, P. and O. Bover (1997): Regional migration in Spain: The effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. *Oxford Bulletin of Economics and Satistics*, 59(2), 215-235.

[4] Basile, R. and M. Causi (2007): Le determinanti dei flussi migratori nelle province italiane: 1991-2000. *Economia e Lavoro*, 2, 139-159.

[5] Basile, R., Girardi, A. and M. Mantuano (2012): Migration and regional unemployment in Italy. *The Open Urban Studies Journal*, 5, 1-13.

[6] Bennett, D. (1998): *Randomness.* Cambridge, Cambridge University Press.

[7] Blanchard, O. and L. Katz (1992): Regional evolutions. *Brookings Papers on Economic Activity*, Economic Studies Program, The Brookings Institution, 23(1), 1-76.

[8] Borts, G. and J. Stein (1964): *Economic Growth in a Free Market.* New York: Columbia University Press.

[9] Brockwell, P. and R. Davis (2003): *Introduction to time series and forecasting.* Berlin, Springer.

[10] Carlstein, E. (1986): The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics*, 14, 1171-1179.

[11] Charemza, W. and D. Deadman (1997): *New directions in econometric practice: General to specific modelling, cointegration, and vector autoregression.* Lyme: Edward Elgar.

[12] Da Vanzo, J. (1978): Does unemployment affect migration? Evidence from microdata. *Review of Economics and Statistics*, 60, 504-514.

[13] Davidson, J. (2000): *Econometric theory.* New York: Wiley.

[14] Davison, A. and D. Hinkley (1997): *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

[15] Faini, R., Galli, G., Gennari, P. and F. Rossi (1997): An empirical puzzle: Falling migration and growing unemployment differentials among Italian regions. *European Economic Review*, 4, 571-579.

[16] Feridun, M. (2004): Does immigration have an impact on economic development and unemployment? Empirical evidence from Finland (1981-2001). *International Journal of Applied Econometrics and Quantitative Studies*, 1, 39-60.

[17] Feridun, M. (2005): Economic impact of immigration on the host country: The case of Norway. *Prague Economic Papers*, 4, 350-362.

[18] Feridun, M. (2007): Immigration, income and unemployment: An application of the bounds testing approach to cointegration. *Journal of Developing Areas*, 41(1), 37-51.

[19] Gibbons, S. and H. Overman (2012): Mostly pointless spatial econometrics?. *Journal of Regional Science*, 52(2), 172-191.

[20] Granger, C. (1969): Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424-438.

[21] Granger, C. (1988): Causality, cointegration, and control. *Journal of Economic Dynamics and Control*, 12, 551-559.

[22] Gross, D. (2004): Impact of immigrant workers on a regional labour market. *Applied Economics Letters*, 11, 405-408.

[23] Hao, B. and W. Zheng (1998): *Applied symbolic dynamics and chaos.* World Scientific: Singapore.

[24] Heckman, J. (1999): Causal parameters and policy analysis in economics: A twentieth century retrospective. *National Bureau of Economic Research*, Working Paper 7333.

[25] Hendry, D. F. (2004): Causality and exogeneity in non-stationary economic time series. In Hall, S. G. (ed.): *New directions in macromodelling* (Contributions to Economic Analysis, Volume 269), 21-48. Bingley: Emerald Group Publishing Limited.

[26] Herrera, M, Ruiz, M, and J. Mur (2013): Detecting dependence between spatial processes. *Spatial Economic Analysis* (forthcoming).

[27] Hoover, K. (2004): Lost causes. *Journal of the History of Economic Thought*, 26, 149-164.

[28] Isard, W. (1971): *Métodos de análisis regional: Una introducción a la ciencia regional.* Barcelona, Ariel.

[29] Isserman, A. M. and J. Merrifield (1982): The use of control groups in evaluating regional economic policy. *Regional Science and Urban Economics,* 12, 43- 58.

[30] Islam, A. (2007): Immigration unemployment relationship: The evidence from Canada. *Australian Economic Papers*, 46(1), 52-66.

[31] Konya, L. (2000): Bivariate causality between immigration and long-term unemployment in Australia, 1981-1998. *Victoria University of Technology* Working Paper, 18/00.

[32] Krugman, P. (1991): Increasing returns and economic geography. *Journal of Political Economy*, 99, pp. 483-499.

[33] Lesage, J. and K. Pace (2009): *Introduction to spatial econometrics.* London, Chapman & Hall/CRC.

[34] López, F., Matilla-García, M., Mur, J., and M. Ruiz Marín (2010): A non-parametric spatial independence test using symbolic entropy. *Regional Science and Urban Economics*, 40, 106-115.

[35] Maddala, G.S. and I.M. Kim (1999): *Unit roots cointegration and structural change.* Cambridge: Cambridge University Press.

[36] Marr, W. and P. Siklos (1994): The link between immigration and unemployment in Canada. *Journal of Policy Modeling*, 16(1), 1-25.

[37] Marr, W. and P. Siklos (1995): Immigration and unemployment: A Canadian macroeconomic perspective. In DeVoretz (ed.), *Diminishing returns: The economics of Canada's recent immigration policy*, 293-330. Toronto: C.D. Howe Institute.

[38] Matilla-García, M. and M. Ruiz Marín (2008): A non-parametric independence test using permutation entropy. *Journal of Econometrics*, 144, 139-155.

[39] Matilla-García, M. and M. Ruiz Marín (2009): Detection of non-linear structure in time series. *Economics Letters*, 105, 1-6.

[40] Pagan, A. (1989): 20 years after. Econometrics 1966:1986. In B. Cornet and H. Tulkens (eds): *Contribution to operations research and econometrics,* pp 53-97 Cambridge: MIT Press

[41] Partridge, M., Boarnet, M., Brakman, S. and G. Ottaviano (2012): Introduction: Whither spatial econometrics?. *Journal of Regional Science*, 52, 167-171.

[42] Pearl, J. (2009): *Causality: Models, reasoning, and inference.* 2nd ed. Cambridge: Cambridge University Press.

[43] Pischke, J. and J. Velling (1997): Employment effects of immigration to Germany: An analysis based on local labor markets. *The Review of Economics and Statistics*, 79(4), 594-604.

[44] Pissarides, C. and I. McMaster (1990): Regional migration, wages and unemployment: Empirical evidence and implications for policy. *Oxford Economic Papers,* Oxford University Press, 42(4), 812-831.

[45] Pissarides, C. and J. Wadsworth (1989): Unemployment and the inter-regional mobility of labour. *Economic Journal*, 99(397), 739-755.

[46] Pope, D. and G. Withers (1993): Do migrants rob jobs? Lessons from the Australian history 1961-1991. *Journal of Economic History*, 53, 719-742.

[47] Ripley, B. (1988): *Statistical Inference for Spatial Processes*. Cambridge, Cambridge Universty Press.

[48] Rohatgi, V. (1976): *An introduction to probability theory and mathematical statistics*. Wiley, New York.

[49] Ruiz, M., López, F. and A. Páez (2009): Testing for spatial association of qualitative data using symbolic dynamics. *Journal of Geographical Systems*, 10.1007/s10109-009-0100-1.

[50] Sims, C. (1972): Money, income and causality. *American Economic Review*, 62, 540-552.

[51] Tian, G. and J. Shan (1999): Do migrants rob jobs? New evidence from Australia. *Australian Economic History Review*, 39(2), 133-142.

[52] Wiener, N. (1956): The theory of prediction. In Beckenbach, E. (ed.): *Modern mathematics for engineers*: Mc Graw: New York.

[53] Zellner, A. (1988): Causality and causal laws in economics. *Journal of Econometrics*, 39,7-21

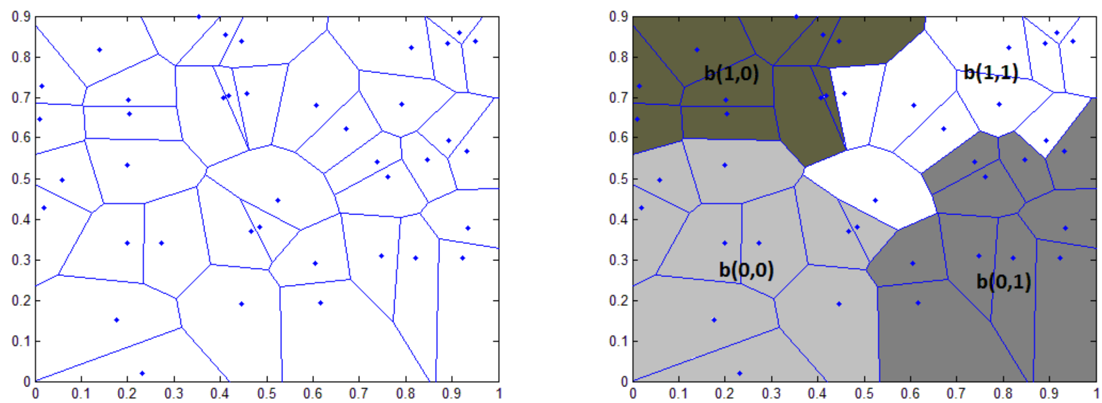
Appendix A: Spatial Block Bootstrapping

We use spatial bootstrapping techniques to test for the null hypothesis of (30). Our purpose is testing for causality from x to y, but preserving the spatial dependence in each series. If there is, in fact, cross-sectional dependence, is crucial that the resampling preserves as much as possible the spatial structure of the data set. Obviously, the simple random resampling is not useful here. Moreover, few results have been established for the case of bootstrapping empirical processes for spatially dependent data (Davison and Hinkley, 1997).

In our experiment, the blocks are formed according to a distance criterion. In the first place, we defined b fixed points. Let us call them the *buoys* of the SBB. Each data point is assigned to the nearest buoy to form the corresponding block (other procedures of block formation can be used). The space has been partitioned into b blocks that now are re-sampled with replacement.

The example of Figure A.1 illustrates our procedure. The original sample space appears on the left panel with 40 data points. We are going to use four blocks, $b = 4$, and the buoys correspond to the four vertices of the rectangle: $\{b\,(0,0)\,, b\,(1,0)\,, b\,(1,1)\,, b\,(0,1)\}$. The same ordering will be used in the SBB. The blocks formed appear on the right panel where each data point has been assigned to the nearest buoy. Then, we apply the resampling with replacement using the four blocks. Let us assume that the result of a given re-sample is $\{b\,(0,0)\,, b\,(1,0)\,, b\,(1,0)\,, b\,(1,1)\}$. This means that, in the bootstrapped sample, the block $b\,(0,0)$ will remain in its current position, the same with the second block $b\,(1,0)$. However, the observations of the second block will be copied, and distributed according to a distance criterion, in the spatial layout of the third block, $b\,(1,1)$; finally, the 10 observations of the third block will be copied and distributed in the spatial layout of the fourth block, $b\,(0,1)$.

Figure A.1: Example of Spatial Block Bootstrapping



(a) Original Sample *(b) Non-Overlapping Blocks*

This bootstrap scheme adapts well to our case because the contiguity criterion used in the experiment is based on the k-nearest neighbors. Changes in the neighborhood criterion would imply changes in the spatial bootstrapping scheme. Overall, our results show that the SBB procedure preserves much of the spatial structure of the original series. Small mismatches appear along the borders of the blocks because there is a non negligible probability that observations have different neighbors in the bootstrapped layout.